note technical note tech

Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe with an Updated Traffic Sample from Washington, DC

Andrew Crowell, FAA ANG-C41

Brian Schnitzer, General Dynamics Information Technology

Mike Paglione, FAA ANG-C41

February 2013

DOT.FAA/TC-TN13/18

U.S. Department of Transportation

Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

NOTICE

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT.FAA/TC-TN13/18		

4. Title and Subtitle Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe with an Updated Traffic Sample from Washington, DC	**5. Report Date** February 2013 **6. Performing Organization Code** ANG-C41
7. Author(s) Andrew Crowell, Brian Schnitzer, and Mike Paglione	**8. Performing Organization Report No.** DOT.FAA/TC-TN13/18
9. Performing Organization Name and Address U. S. Department of Transportation Federal Aviation Administration, William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Agency Name and Address U. S. Department of Transportation NextGen Implementation and Integration Office Washington, D. C. 20590	**13. Type of Report and Period Covered** Technical Note **14. Sponsoring Agency Code** DOT

15. Supplementary Notes

The authors identified above represent the following organizations: Andrew Crowell and Mike Paglione with FAA's NextGen Concept Analysis Branch (ANG-C41) and Brian Schnitzer with General Dynamics Information Technology.

16. Abstract

The Federal Aviation Administration (FAA) is currently implementing a number of improvements to the National Airspace System (NAS) in the United States under a multi-agency initiative called the Next Generation Air Transportation System (NextGen) Program. The Separation Management and Modern Procedures Project is a NextGen initiative and its objective is to implement the En Route Automation Modernization (ERAM) strategic conflict probe on the radar controller display utilizing ERAM's Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems. The FAA's Air Traffic Organization's En Route Program Office (ATO-E) has employed the FAA's Concept Analysis Branch (ANG-C41) to conduct a series of independent evaluations on prototype enhancements to the TM and CP sub-systems and has contracted the prime contractor of ERAM, Lockheed Martin, under FAA Task Orders 45 and 51 to develop these prototypes within the ERAM architecture. This paper describes the fourth and final in a series of integrated experiments to study these enhancements. The experiment consists of simulated runs using the ERAM system with various combinations of prototypes enabled and with various parameter settings. The trajectory modeling and conflict probe performance of these treatment runs are compared against a baseline run which represents the current state of the live ERAM CP. Recorded data from real flights in the Washington Center (ZDC) was processed to create a realistic air traffic scenario sample. This technical note provides a detailed description of the analyses performed as well as the results of these analyses.

17. Key Words Separation Management, Conflict Probe, En Route Automation Modernization, ERAM, Trajectory Modeling	**18. Distribution Statement** This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	40	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Executive Summary

The Federal Aviation Administration (FAA) is currently implementing a number of improvements to the National Airspace System (NAS) in the United States under a multi-agency initiative called the Next Generation Air Transportation System (NextGen) Program. The NextGen operational concept envisions a future air traffic environment managed by aircraft trajectory with advances in ground automation like the conflict probe. The Separation Management and Modern Procedures Project is one of these NextGen initiatives and its objective is to implement the En Route Automation Modernization (ERAM) strategic conflict probe on the radar controller display. The strategic conflict probe utilizes ERAM's Trajectory Modeling (TM) and Conflict Probe (CP) sub-systems to notify air traffic controllers when aircraft will violate separation standards as much as 20 minutes in the future. The FAA's Air Traffic Organization's En Route Program Office (ATO-E) contracted the prime contractor of ERAM, Lockheed Martin, under FAA Task Orders 45 and 51 to develop these prototypes within the ERAM architecture so the FAA may evaluate their efficacy. ATO-E has employed the FAA's Concept Analysis Branch (ANG-C41) to conduct a series of independent evaluations on performance enhancements to the TM and CP sub-systems.

This paper describes the fourth in a series of integrated experiments to study these enhancements. The experiment consists of simulated runs using the ERAM system with different combinations of prototypes enabled and with various parameter settings. The TM and CP performance of these treatment runs are compared to that of the baseline run, which represents the current state of the live ERAM system. Each of these runs is based on the same scenario, which is generated by time-shifting real traffic data recordings to induce conflicts. The traffic data is from a 2010 recording of the Washington Center during peak hours. This is the third and final scenario in a series to be analyzed with this approach.

This experiment expands upon the analysis done in three previous experiments, by focusing in on certain factors and applying the experiment to a third scenario. This is the last experiment to be performed in this series, so final conclusions and recommendations can now be made. Prior to this experiment, the recommendations were to set the lateral conformance bound to 1.0 nm and the longitudinal to 1.25 nm with Growth Adherence Bounds (GAB) and Conflict Geometry Separation (CGS) enabled. The recommendation for lateral conformance bounds was confident enough to be considered a final recommendation, but the others needed to be confirmed. This experiment attempted to confirm the GAB and CGS recommendations and to expand upon the longitudinal recommendation to investigate.

In this experiment, the GAB prototype showed as much as a 10% improvement to False Alerts (FA) without any impact on Late Alerts (LA), Missed Alerts (MA), or warning time (WT). This is consistent with previous experiments, and as a result, it is the final recommendation of the Concept Analysis Branch that this prototype be pursued for addition into the live ERAM system.

The CGS prototype also showed as much as a 10% reduction in FAs. It did, however, have a small impact on LAs and MAs, increasing them by up to 50%. It is important to note, however, that, because of the small sample size, a 50% increase represents only two to three alerts. Although the CGS prototype on its own may increase the LAs and MAs, combined the with FA32 enhancements, it shows no impact to the LAs and MAs at the recommended settings. The major negative impact of the CGS prototype is now the WT. It may reduce the WT by up to 19%. However, even with this 19% reduction, the WT is still well above the 180 second requirement. Though CGS had to undergo many modifications to reach its current performance levels, the

results observed here are consistent with previous experiments, and so it is the final recommendation of the Concept Analysis Branch that this prototype be pursued for addition into the live ERAM system.

For this experiment, the longitudinal conformance bound was split into its two parts: track monitoring (TM) and conflict detection (CD). This was done in order to determine if any additional performance gains could be observed by modifying the two separately. It was determined that modifying the TM bound by itself has almost no impact to any of the factors, though setting it lower than the current recommendation of 1.25 nm may actually negatively impact performance. The CD bounds have a much greater impact to the probe, with the possibility of reducing the FAs by up to 9% with no significant impact to LAs, MAs, or WT. Once again, these findings are consistent with previous experiments, so it is the final recommendation of the Concept Analysis Branch that the longitudinal conformance bounds be set to 1.25 nm for the TM and CD portions.

Final recommendations for the settings and prototypes of ERAM

Prototypes	Lateral	Longitudinal	Likelihood
FA32 ap1; GAB; CGS	1.0	1.25	4|8|20

The table above shows the final recommendations being made for all prototypes and parameter changes studied in all five experiments conducted for Separation Management Functional Areas 18 and 32 in calendar years 2010 through 2012. This table includes the recommendations of the likelihood experiment being performed concurrently. With these settings, for this scenario, the False Alerts were reduced by 55.5% and 25th percentile of warning time was reduced by 19%. Missed Alerts and Late Alerts were not affected at all.

Table of Contents

List of Figures

List of Tables

1 Introduction

The Federal Aviation Administration (FAA) currently has many projects underway for improving the National Airspace System (NAS) that fall into the realm of the Next Generation Air Transportation System (NextGen). Separation Management: Modern Procedures is concerned with the performance and usability of the strategic Trajectory Predictor (TP) and Conflict Probe (CP) of the En Route Automation Modernization (ERAM). The current goal is to improve the performance of the strategic CP by reducing the nuisance alerts to acceptable levels, without adversely affecting its performance on correct alerts. This technical note details a study performed by the Concept Analysis Branch of the FAA in support of this goal.

1.1 Background to Study

In 2011 the FAA's Concept Analysis Branch (ANG-C41) published two reports of integrated experiments that were performed on a single day of recorded, time-shifted air traffic data from a single Air Route Traffic Control Center (ARTCC, Center) [Crowell et al, December 2011a] [Crowell et al, December 2011b]. The recording date was March 17, 2005, and included traffic in the Washington, D.C. (ZDC) ARTCC.

The documents reported that the lateral conformance bound being used in the current live system (2.5 nm) is inefficient and much larger than it needs to be which results in a generation of nuisance alerts. A recommendation was made to greatly reduce this bound, possibly to as low as 1.0 nm, and even lower once ADS-B is more prevalent in the NAS.

Longitudinal conformance bound was found to be much closer to a preferred value. The current value is set to 1.5 nm, and a recommendation was made to set it to 1.25 nm.

The likelihood threshold function was determined to be used inefficiently in the current system with a mapping of 10|20 (0.0 likelihood alerted at 10 minutes, 1.0 likelihood alerted at 20 minutes). It was recommended that these values change, but no value could be recommended at the time. Instead a future study was planned, which will be performed and published in 2012.

Finally, the three prototype algorithms, Forced Trajectory Rebuild (FTR), Growth Adherence Bounds (GAB), and Conflict Geometry Separation (CGS), were studied. Only GAB was recommended for addition into the probe. FTR and CGS both showed improvement in certain circumstances but overall hindered the CP more than they helped it. Additional study was recommended for FTR and CGS.

In 2012 another study was performed using similar methods on a new scenario [Crowell et al, December 2012a]. The scenario contained flights from Chicago Center (ZAU) on February 11, 2010. This document recommended settings of 1.0 nm for lateral conformance bound and 1.25 nm for longitudinal conformance bound. It also made a recommendation to postpone any further research on the FTR prototype, but continue research on GAB and CGS. GAB showed improvements in both studies and was a strong candidate for addition to the CP. However, CGS still showed mixed results and more information was needed to make a final recommendation.

In parallel with this study, another experiment, Experiment 5, is being performed [Crowell et al, December 2012b]. This study focuses specifically on the settings of the likelihood threshold. Recommendations of Experiment 5 were to use the 4|8|20 setting for the live system, as this provided a significant improvement to CP performance.

This experiment, denoted as Experiment 4, draws from and builds upon the recommendations from previous and parallel experiments. The two remaining prototypes are the main focus of this experiment, but particularly the CGS prototype, due to its mixed results in previous experiments. This experiment is expected to confirm previous conclusions about the GAB prototype and provide a final conclusion for the CGS prototype. The previous recommendation of 1.0 nm lateral conformance bound is used as a constant in this experiment. The longitudinal conformance bound is studied in more detail in order to provide a better understanding of its effects on CP performance. Finally, the FTR prototype is not included in this experiment due to previous recommendations. The 4|8|20 likelihood threshold parameter is also used as a constant, as recommended from previous experiments.

The study uses a traffic sample from the Washington Center (ZDC). The traffic was recorded on April 30, 2010 and contains 2734 flights with 239 conflicts.

1.1.1 Prototype Enhancements

This study analyzes only two of the three prototype algorithms studied in previous experiments [Crowell et al, December 2011b] [Crowell et al, December 2012a]. The algorithms evaluated are Growth Adherence Bounds and Conflict Geometric Separation. Each is briefly defined in the following sections and a complete description of the algorithms can be found in [Lapihuska, November 2011].

1.1.1.1 Growth Adherence Bounds

The approach for the prototype Growth Adherence Bounds (GAB) algorithm is to perform a filter on the standard conformance (adherence) bounds that will modify the conformance bounds as the probe traverses temporally through the route. The algorithm will apply smaller conflict detection adherence bounds to near-in time segments in lieu of the standard conformance bounds currently used. The bounds are gradually increased as the probe proceeds further down the predicted route path until the graduated bounds reach the same size as the standard bounds.

The reasoning supporting the algorithm is the fact that flights typically only deviate gradually from the predicted path, so that near-in time segments can have smaller conformance bounds than the bounds used further down the route.

The GAB is most effective when the “age” of the trajectory (time between trajectory build start time and current time) is small – that is, when the trajectory has recently been updated. GAB is also more effective when the flights are diverging over time and if the period of conflict is limited to the next few minutes. A significant contribution of GAB is to shorten the duration of notification after minimum separation has passed and a separation of 6.2 nm has been achieved. The threat of conflict no longer exists even though the flight separation is still within conformance bound distances. Any reduction in notification time reduces controller distraction, so early elimination of conflicts that have already passed critical separation times improves nuisance rate performance.

The GAB design is conceptually based on an earlier MITRE effort [Rosen, 2008] [Bolczak, 2010] designated as “tactical check” and proposed as a NextGen Separation Management enhancement. However, there are significant differences between the two. The prototype GAB applies the growth in both lateral and longitudinal directions while the MITRE approach applied lateral only. The MITRE approach ensured, within the algorithm design, regular trajectory updates – the prototype does not. The MITRE approach applies some asymmetric lateral adaptations based on relationship of the track to the filed route. The prototype applies a symmetric growth factor with respect to the current trajectory.

Using a single lateral conformance bound for all times results in increased nuisance alerts. The GAB algorithm may be especially useful for reducing nuisance alerts in predicting a near-term conflict.

1.1.1.2 Conflict Geometric Separation

For a few specific potential conflict cases, additional processing, called Conflict Geometry Separation (CGS), will be executed. CGS examines the conflict geometry to determine whether or not a conflict should be discounted.

CGS processing will depend on the category of the specific conflict geometry. The three geometry categories are in-trail, parallel, and crossing. An in-trail conflict is a conflict that can occur on a shared segment of a route that is common between two flights and where the two aircraft are flying generally in the same direction. However, in-trail conflicts can also occur between aircraft that do not have common route segments. A parallel conflict occurs when the corresponding trajectory paths and route paths are greater than 6nm from each other and the closure angle for any of the segment corresponding pairs does not exceed 15 degrees. All remaining conflicts that do not meet the definition of either an in-trail or a parallel are defined as a crossing conflict. After categorizing the geometries of the conflicts, specific criteria are examined to determine subsequent action. The CGS algorithm is applied selectively based on the encounter geometries of the conflict flight pair.

In Experiment 2, the algorithm parameters were set to delay the alert until it had a predicted time-to-conflict of just over three minutes. However, this assumed that the closure rate would not increase over those three minutes. In many cases the closure rate did increase, reducing the time-to-conflict to less than three minutes, and in turn causing CGS to generate a Late Alert according to the definition of Late Alerts in [Crowell et al, December 2011a]. In this study, the time-to-conflict was increased to four minutes to try to avoid this situation.

1.1.2 Conflict Probe Parameters

The last two settings manipulated in the treatment runs of this experiment are parameters of the ERAM Conflict Probe. These parameters can be varied independently and affect the probe in different ways. The two parameters changed were longitudinal conflict detection (CD) conformance bound and longitudinal track monitoring (TM) conformance bound.

1.1.2.1 Conformance Bounds

The conformance bounds serve two purposes in ERAM. They determine when a trajectory is built for re-adherence purposes, and they determine when a conflict prediction is made based on a trajectory. The lateral conformance bound is added to the left and right side of the trajectory or flight, whereas the longitudinal conformance bound is added to front and back.

For re-adherence purposes, the conformance bounds create a box that is twice the width of the lateral and twice the length of the longitudinal. It is then placed centered on the predicted position of the flight as shown in Figure 1. If the actual position of the flight is outside of this box, then a new trajectory will be rebuilt to re-adhere to the position of the flight.

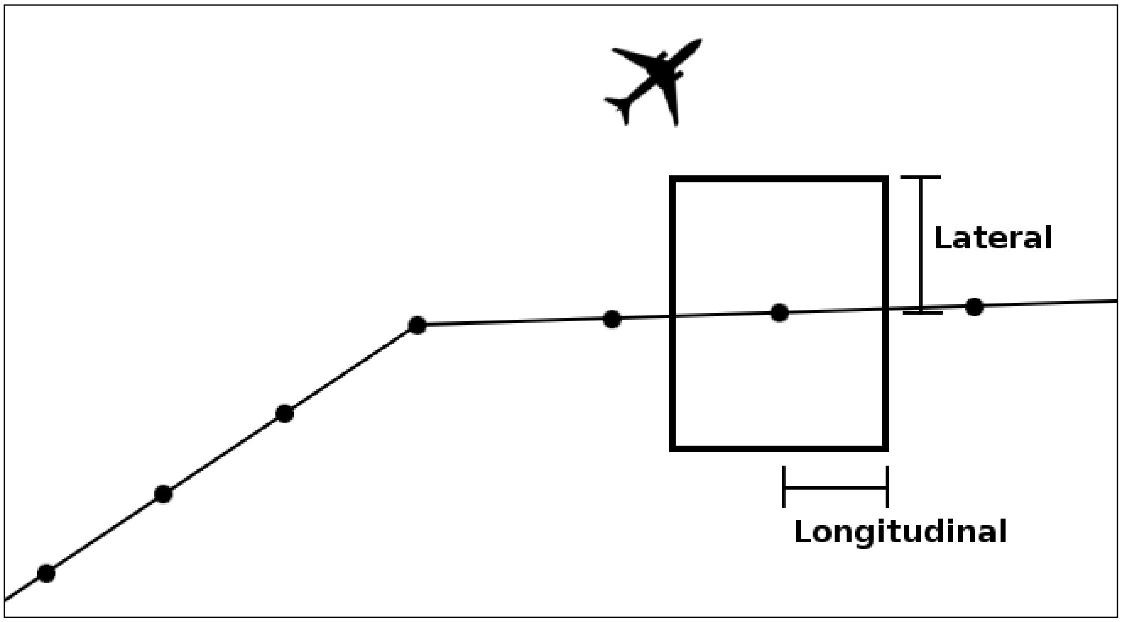

Figure 1. Conformance Bounds Used for Re-Adherence of Trajectory to the Flight Path

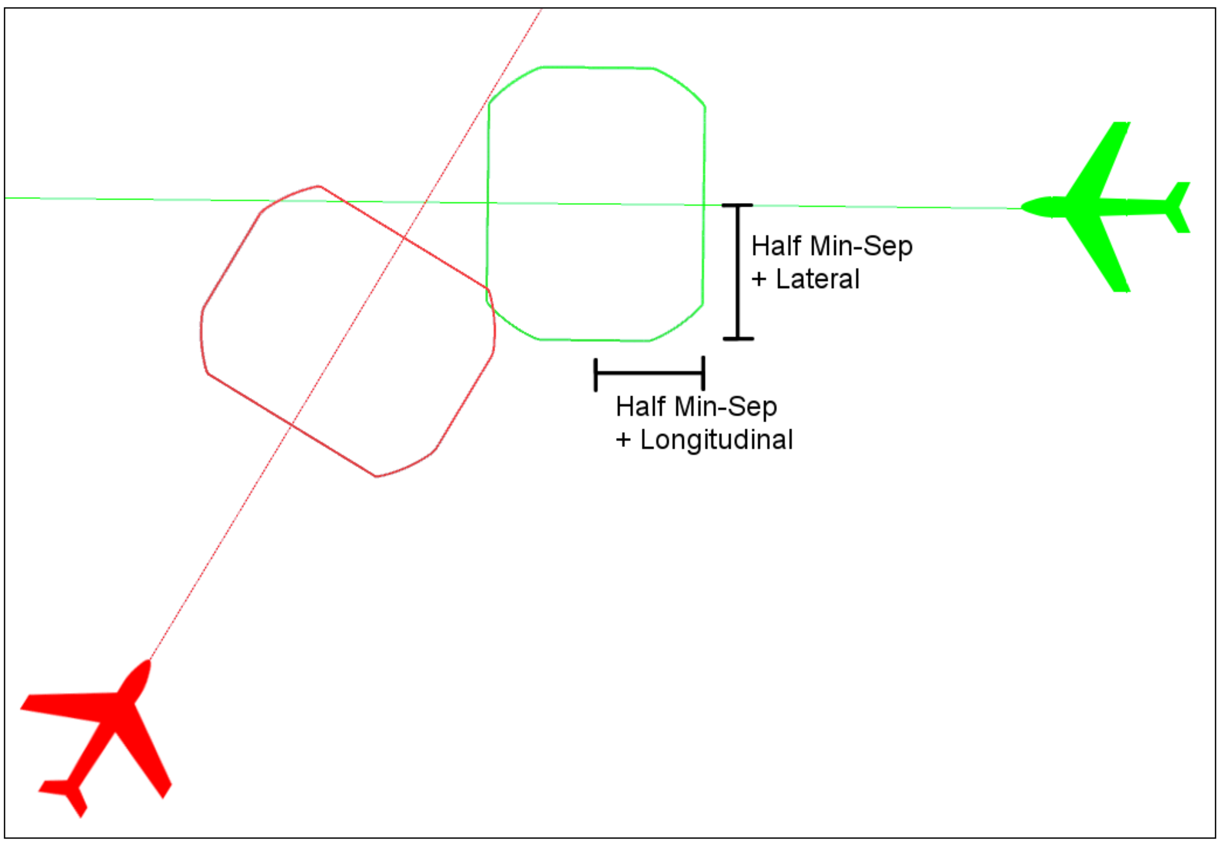

Figure 2. Depiction of Conformance Bounds Used for the Conflict Probe

The use of conformance bounds for the Conflict Probe is much more complex. Several levels of filters are used initially. Then, an octagonal shape is formed using the geometry and conformance bounds of each of the aircraft. This document will not go into the details of this algorithm, but the same results can be visualized by adding the conformance bounds to the required separation of each aircraft to create a box around the predicted position of each aircraft and then cutting off the corners of the boxes with a circular filter. The resulting shape is shown in Figure 2. If these two boxes intersect each other, then a conflict prediction is made.

1.1.3 Previous Work

This study, designated Experiment 4, is the last of two experiments performed in 2012 with the purpose of analyzing the impact of prototype algorithms and parameter changes in ERAM on CP performance. Both studies are a follow up to Experiments 1 and 2 performed in 2011. Experiment 1 was performed to determine if there was a set of parameter adjustments that could be made to the conflict probe in order to improve performance. Experiment 2 analyzed each of the three prototype enhancements to determine if any of them provide a significant improvement to performance. Experiment 3 was a combination of the first two experiments using a new traffic sample from Chicago Center.

1.2 Scope of Study

This document reports on the results of an experiment limited to one six-hour traffic sample collected on April 30, 2010 from the Washington Air Route Traffic Control Center (ZDC). To induce conflicts between aircraft and for evaluation purposes only, the data sample was time-shifted using the methodology documented in [Paglione, 2003].

This experiment follows similar experiment from 2011 and earlier in 2012 that used an older sample of traffic from ZDC recorded in May 2005 and a sample of traffic from Chicago Center (ZAU) recorded in February 2010. This experiment is intended to expand the findings of the previous experiments and make final recommendations.

All of the analyses in this document were performed on a time-shifted scenario. Currently, the metrics available for analyzing performance require a time-shifted scenario to be used in order to generate actual loss of separation that would not occur under normal circumstances. This time-shifting can create some events that the conflict probe will never encounter in a live system. As a result, the reader should be careful not to take any numbers presented in this document out of context. All numbers presented in this document should be used only for comparison to other numbers included in this document, unless otherwise noted. The False Alert, Late Alert, and Missed Alert rates, as well as the warning time values presented in this document do not reflect the actual values of the live ERAM system and should not be considered as such. Because of this, most of the values presented in this document are in the form of percentage change from the baseline results. Though some raw numbers may be presented, they should be considered only in the context of this document.

1.3 Document Organization

This technical note is organized in the following sections: Section 1.1.1 provides a high-level description of the two prototype enhancements being analyzed in this study. Section 2 defines the experiment performed and describes the development of the model along with the final statistical qualities of the model. Section 3 describes the analyses that were performed to evaluate the Conflict Probe (CP) performance. Finally, Section 4 wraps up the conclusions of the performance analyses and makes recommendations based on the findings.

2 Description of Experiment

One of the most powerful inferential statistical approaches is the design, implementation, and synthesis of experiments. Experiments are performed by most researchers and scientists in practically all disciplines. An input stimulus is entered into a process with a set of controllable factors. The uncontrollable factors are not easily manipulated, but through experimental design techniques such as blocking and randomization can be removed from the experiment. The output response variables are the dependent variables of the experiment. They are often determined by application of a metric or measured by a sensor device.

Table 1. Processing Steps for the Experimental Analysis

Step	Description	Section
1 – Problem Definition	Define the problem statement	2.1
2 – Design of Experiment	Design the experiment – The factors, levels of the factors, response variables to be run, and the model to be used for analysis are defined.	2.2
3 – Execute Experiment	Execute the experiment and prepare output data – The system is configured for the experimental runs defined by the design, runs executed, and resulting output data is processed for input into model	3
4 – Implement Model	Implement statistical model defined by the experiment.	3.1.2
5 – Model Results	Examine the results of the model and discuss factor effects	3.1.2 & 3.1.3
6 – Synthesize Impact	Synthesize overall results from the model and publish conclusions.	4

There are many purposes for performing an experiment. For this study, the objective of designing and executing an experiment is to establish (1) which pre-determined factors and interactions of these factors show a statistically significant effect on the ERAM system's performance, and (2) the relative sizes of the determined significant effects. From designing the experiment to concluding on its results, a series of processing steps should be performed as identified in Table 1. The first two steps presented in Table 1 are described in this section, which documents the plan for the experimental analysis. The last four steps are described in Section 3 and Section 4, which present the results by documenting the actual execution and analysis of the experiment.

The integrated experiment used in this study is modified from that used in Experiment 3, based on the knowledge gained from the previous experiments. The purpose of this experiment is to determine if any of the manipulated factors provide a statistically significant improvement to the performance of the Conflict Probe. In order to evaluate this, it is also necessary to determine how each of the factors interacts with one another.

The factors for the prototype enhancements are binary and indicate whether that particular prototype enhancement is on or off. Given the two binary factors and the two continuous factors, the total number of runs required for a full factorial design (assuming three samplings of the continuous functions) would be 36. Since each run must be performed using the live ERAM system in a simulation environment, it is necessary to reduce this number considerably. The experiment was designed using the JMP® software tool and is described in the following sections.

2.1 Definition of the Problem Statement

It must be determined if any of the prototypes or parameter changes can provide a significant improvement to Conflict Probe (CP) performance. CP performance is measured in False Alert, Late Alert, and warning time performance, all of which can vary separately. Low False Alerts, low Late Alerts, and high warning time are the desired qualities of CP performance. The two prototypes covered in this study are intended to improve False Alert performance. A significant improvement to CP performance will be recognized if a prototype significantly improves False Alert performance, and does not significantly degrade Late Alert performance. It is also desirable to avoid degrading warning time performance, but this is not a requirement in order for a CP performance improvement to be recognized. For this study, the problem statement is expressed as follows:

> *Through a set of purposeful runs of ERAM, input with a ZDC time-shifted test traffic scenario, the experiment shall determine the statistically significant impact that the Growth Adherence Bounds or Conflict Geometric Separation prototype algorithms, or longitudinal CD or TM adherence have in terms of trajectory and conflict prediction accuracy performance.*

A significant change, whether it is improvement or degradation, is defined as a change in the respective metric (False Alerts, Late Alerts, or warning time) that is greater than the confidence intervals of the statistical model. These confidence intervals are discussed in the next section.

2.2 Design of Experiment

In order to reduce the number of runs required to perform the analysis, a d-optimal design was used rather than a full factorial [NIST/SEMATECH, 2011]. A d-optimal design can be thought of as selecting the corner points on a six-dimensional hypercube created from the six factors, allowing the model to interpolate in between these corner points. For the two continuous factors, center points are also selected in those dimensions allowing a quadratic interpolation to be performed instead of just a linear interpolation.

2.2.1.1 Factors

The factors used in the experiment included settings of ERAM that can be changed in the current version as well as prototype upgrades. The prototype upgrades would require code enhancements to the current version of ERAM.

The longitudinal conflict detection (CD) and track monitoring (TM) conformance bounds were varied independently from each other. This variance used the prototype changes in FA18 Interim 2 [Crowell et al, June 2011] [Lapihuska, 2011] that decoupled the TM bounds from the CD bounds. These bounds are continuous factors, modeled using a quadratic equation. Ranges of the two continuous factors are listed in Table 2.

Table 2. Continuous Factors of the Integrated Experiment

Factor	Min	Max
Longitudinal Conflict Detection Bound	1.0 nm	1.5 nm
Longitudinal Track Monitoring Bound	0.5 nm	1.0 nm

The prototype enhancements are binary factors, either running or not running. These enhancements were described in Section 1.1.1. They include GAB (Growth Adherence Bounds) and CGS (Conflict Geometric Separation).

The settings described above resulted in the 20 runs shown in Table 3. Not shown in this table are the settings used currently in the deployed version of ERAM. This run is referred to as the baseline (BL) run.

Table 3. Runs for the integrated experiment

Run	GAB	CGS	Lon TM	Lon CD
1	Off	Off	0.50	1.00
2	Off	Off	0.50	1.25
3	Off	Off	0.75	1.50
4	Off	Off	1.00	1.00
5	Off	On	0.50	1.50
6	Off	On	0.75	1.00
7	Off	On	1.00	1.25
8	Off	On	1.00	1.50
9	On	Off	0.50	1.50
10	On	Off	0.75	1.00
11	On	Off	1.00	1.50
12	On	On	0.50	1.00
13	On	On	0.50	1.50
14	On	On	0.75	1.25
15	On	On	1.00	1.00
16	On	On	1.00	1.50
17	Off	Off	1.25	1.25
18	Off	On	1.25	1.25
19	On	Off	1.25	1.25
20	On	On	1.25	1.25

There are several constants that are used in all 20 treatment runs. Some of these constants differ from the BL run, which is why the BL is not shown in the above table. The constants used for the treatment runs are shown in Table 4.

Table 4. Constant settings used for all treatment runs.

FTR	Lat TM	Lat CD	Likelihood
Off	1.0	1.0	4|8|20

The BL run differs in some of these constants, as well as some of the variables. Due to these differences, the baseline run cannot be used within the experimental model. In this experiment the BL is only used for comparison to treatment runs and all percentage differences are generated as differences from the BL. Table 5 shows what the settings were for the BL.

Table 5. Settings of the baseline run (BL) used for this experiment.

FTR	GAB	CGS	Lat TM	Lat CD	Lon TM	Lon CD	Likelihood
Off	Off	Off	2.5	2.5	1.5	1.5	10|20

2.2.1.2 Model

The initial model allowed both of the continuous factors to have at most a quadratic effect. It was assumed that all factors could interact only in pairs (two-way interactions only). The constant or overall mean effect is represented in the model as μ, and $\varepsilon_{n(fghijk)}$ represents the assumption of independently normally distributed random error with a zero mean. All factors are assumed to be additive. The model is defined as in Eq. 1.

Response:

$$
\begin{aligned}
R^0_{fghijk} &= \mu + LonTM_h^2 + LonTM_h + LonCD_i^2 + LonCD_i \\
&+ LonTM_h LonCD_i + GAB_f LonTM_h + GAB_f LonCD_i \\
&+ CGS_g LonTM_h + CGS_g LonCD_i + GAB_f CGS_g + GAB_f \\
&+ CGS_g + \varepsilon_{n(fghi)}
\end{aligned}
\qquad \textbf{Eq. 1}
$$

Where:
GAB_f = growth adherence bounds prototype, f = on, off
CGS_g = conflict geometric separation prototype, g = on, off
$LonTM_h$ = longitudinal TM conformance bounds in nautical miles, h = 0.5, 0.75, 1.0
$LonCD_i$ = longitudinal CD conformance bounds in nautical miles, i = 1.0, 1.25, 1.5
$\varepsilon_{n(fghi)}$ = random error, n = 1, 2, … for all f, g, h, i

This model fits the data very well for the FA % response (Figure 3) and fairly well for the WT % response (Figure 5). The LA % response does not fit as well (Figure 4), but this is expected due to the low sample size for LA %. The horizontal blue line in these figures represents the mean value of the samples and the red curves indicate the 95% confidence interval. The significance is quickly established in a leverage plot by determining if the confidence interval intersects the mean. No intersection indicates insignificance. All responses demonstrate statistical significance.

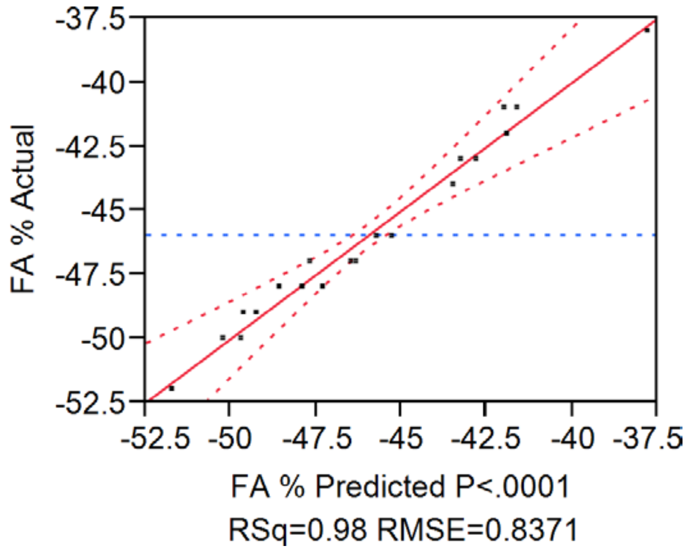

Figure 3. FA % Response Model Fit to Data

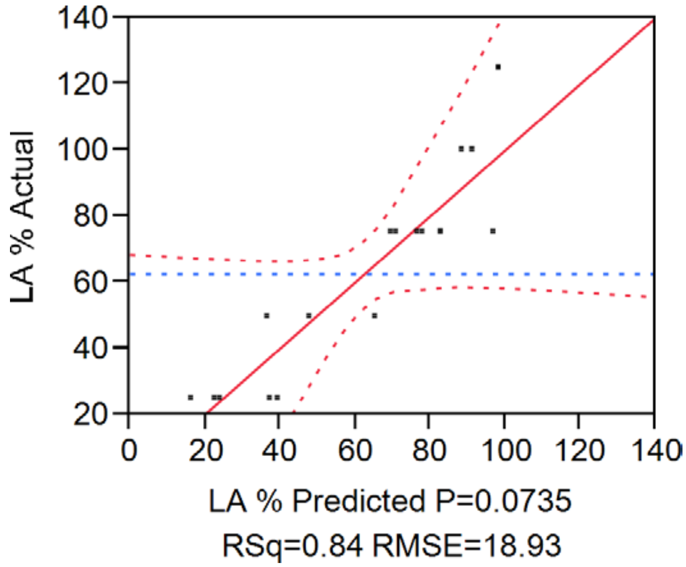

Figure 4. LA % Response Model Fit to Data

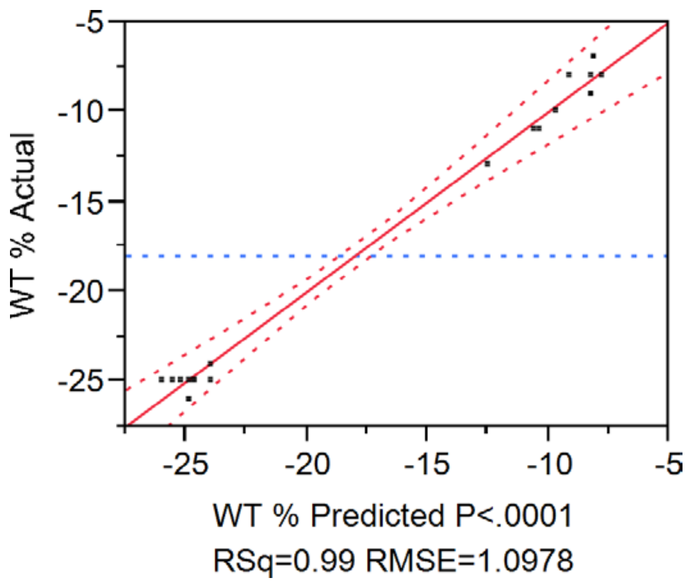

Figure 5. WT % Response Model Fit to Data

The model also relies on the assumption that the random error $\varepsilon_{n(fghi)}$ is normally distributed. The residual errors should therefore be tested for normality. Figure 6 shows histograms and normal quantile plots for the responses. The normal probability plots illustrate that for each response, the model errors fall within the confidence interval along the diagonal line of the plot, indicating that each residual is at least approximately normally distributed. This provides evidence that the residual errors are normal and the model is indeed appropriate.

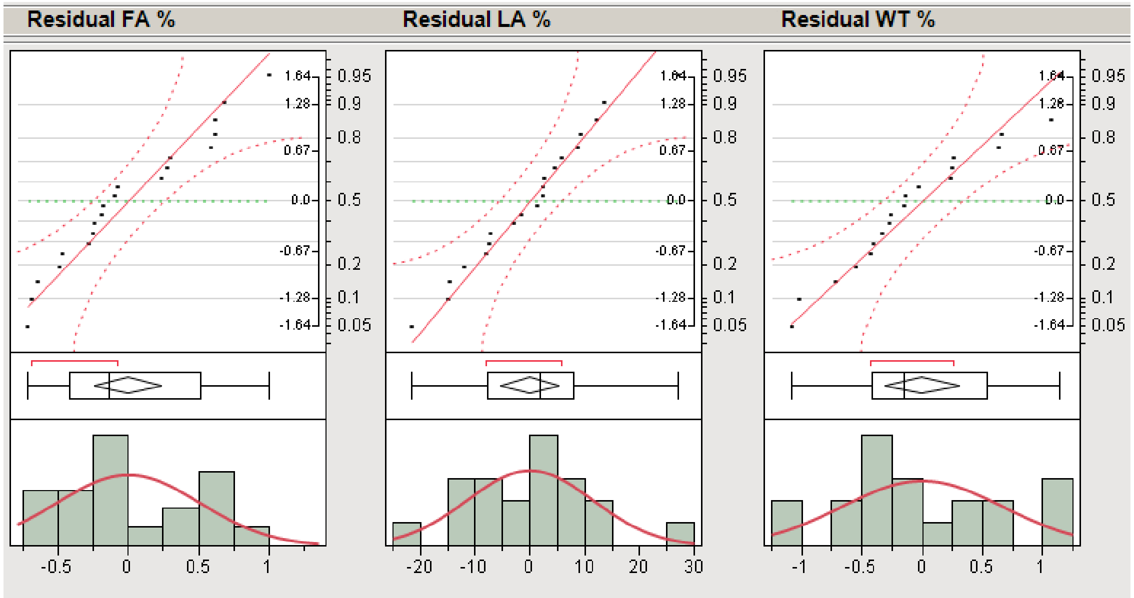

Figure 6. Residual Error Plots for FA, LA, and WT

3 Performance Evaluation

The performance evaluation analyses used in this study are similar to those used in Experiment 2 [Crowell et al, December 2011b]. The metrics used are those described in the documentation of Experiment 1 [Crowell et al, December 2011a]. An integrated experiment was designed, similar to that used in Experiment 2 but containing different settings for the likelihood and longitudinal parameters, as described in Section 2.2.1.1. Unlike Experiments 1 and 2, this experiment was designed to allow the analysts to determine the effects of prototype enhancements and parameter changes in a single experiment. The analysis on the conflict probe (CP) performance is described in detail.

3.1 Conflict Probe Analysis

Several analyses were performed on this experiment to determine the effects that each factor has on the performance of the Conflict Probe (CP). Since this experiment focuses on both the prototype enhancements and the parameter changes, the null hypothesis is as follows:

> *A significant Conflict Probe performance improvement is not observed through parameter changes of the longitudinal conflict detection or track monitoring conformance bounds, nor through the prototype enhancements of Growth Adherence Bounds or Conflict Geometry Separation.*

A significant performance improvement is defined as a reduction in False Alert Rate greater than the confidence interval, no increase in Late Alert Rate greater than the confidence interval, and a 25^{th} percentile of warning time above the three minute threshold. All of these requirements must be true in order for it to be considered a significant improvement and to reject this null hypothesis.

The study documented here will attempt to reject this null hypothesis, therefore showing that these enhancements or parameter changes do indeed provide a significant improvement to the ERAM system.

In all analyses performed in this study the baseline (BL) used contains all of the settings of the current live system, plus the trajectory lateral modeling enhancements defined under Function Area 32 ap1 [McKay, 2011]. The result of the analysis on the FA32 ap1 was a recommendation for addition of those enhancements to the trajectory modeler. As a result, all following experiments assume those lateral modeling enhancements will be included in the system when these additional conflict probe enhancements are implemented.

Table 6 shows the alert type counts for the 20 treatment runs and the baseline. This table is mainly here for documentation of the experiment, but a few observations can be made from it. First, there is very little difference between the MA and LA counts from run to run, so no significant impacts to these responses are expected to be observed. Second, all treatment runs reduce the FA count by nearly 50% or more. These two observations together indicate that all of these settings would greatly improve the performance of the CP.

Table 6. Alert Type Counts

Run	GAB	CGS	Lon TM	Lon CD	VA	FA	MA	LA
1	Off	Off	0.50	1.00	223	656	3	2
2	Off	Off	0.50	1.25	224	702	4	1
3	Off	Off	0.75	1.50	223	760	3	2
4	Off	Off	1.00	1.00	220	639	4	2
5	Off	On	0.50	1.50	216	715	5	1
6	Off	On	0.75	1.00	216	621	3	2
7	Off	On	1.00	1.25	214	633	5	2
8	Off	On	1.00	1.50	214	685	5	2
9	On	Off	0.50	1.50	223	694	3	2
10	On	Off	0.75	1.00	220	636	4	3
11	On	Off	1.00	1.50	221	692	3	3
12	On	On	0.50	1.00	214	613	5	2
13	On	On	0.50	1.50	215	644	5	2
14	On	On	0.75	1.25	212	619	3	5
15	On	On	1.00	1.00	212	588	5	3
16	On	On	1.00	1.50	214	643	4	3
17	Off	Off	1.25	1.25	225	714	4	2
18	Off	On	1.25	1.25	211	638	6	3
19	On	Off	1.25	1.25	220	653	5	2
20	On	On	1.25	1.25	213	606	4	3
BL	Off	Off	1.50	1.50	225	1217	3	1
IBL	Off	Off	1.50	1.50	230	1363	4	3

As always, it is important to consider the warning time of the alerts. There are two metrics used to evaluate the warning time. The first is Adjusted LA, shown in Table 7. This metric adjusts the LA+MA count by the amount of warning time provided by the conflict probe. This will reduce the value for a LA that has a high warning time to some value between 1.0 and 0.0, whereas a MA with no warning time will count as a value of 1.0. This metric provides a much better view of the differences between the BL and treatment runs. Although the LA+MA count is not that different in many of the treatments, the Adjusted LA shows that the alerts in the baseline have generally a higher warning time. Run 2 provides the best Adjusted LA of the treatment runs, with an increase of less than one over the BL.

Table 7. Adjusted Late Alert Value Compared to LA+MA Counts

Run	GAB	CGS	Lon TM	Lon CD	LA+MA	Adj LA
1	Off	Off	0.50	1.00	5	4.31
2	Off	Off	0.50	1.25	5	3.37
3	Off	Off	0.75	1.50	5	4.26
4	Off	Off	1.00	1.00	6	4.91
5	Off	On	0.50	1.50	6	3.97
6	Off	On	0.75	1.00	5	4.26
7	Off	On	1.00	1.25	7	5.29
8	Off	On	1.00	1.50	7	5.29
9	On	Off	0.50	1.50	5	4.31
10	On	Off	0.75	1.00	7	5.77
11	On	Off	1.00	1.50	6	5.42
12	On	On	0.50	1.00	7	5.30
13	On	On	0.50	1.50	7	5.67
14	On	On	0.75	1.25	8	7.17
15	On	On	1.00	1.00	8	6.37
16	On	On	1.00	1.50	7	5.83
17	Off	Off	1.25	1.25	6	5.19
18	Off	On	1.25	1.25	9	6.09
19	On	Off	1.25	1.25	7	3.91
20	On	On	1.25	1.25	7	4.97
BL	Off	Off	1.50	1.50	4	2.56
IBL	Off	Off	1.50	1.50	7	6.30

Table 8 shows the warning time metrics for each of the treatment runs and the baselines. Three warning time metrics are used. The median is included to give an idea of how the run performed overall. The 25th percentile is the metric of most interest which illustrates how the lower end of the alerts performed with regard to warning time. This metric is used because it represents how close the lower end of warning time distribution is to being called Late Alerts. Increasing this value is much more desirable than increasing the median, which is often far above the warning time requirement of 180 seconds. The inter-quartile range (IQR) illustrates the range between the 75th percentile and the 25th percentile. This value can help explain some of the differences between runs. A larger IQR indicates that a conflict probe increases the warning time of those alerts that already have a lot of warning time, or decreases the warning time of those that have little. The IQR can be observed along with the 25th percentile value to get an idea of the shape of the curve of warning times. From here on, the 25th percentile of warning time will be used as the main warning time metric.

This table shows that overall the runs performed very well in regards to warning time. The lowest 25th percentile is 255 seconds, which is still 75 seconds above the required 180 seconds, and only 87.5 seconds lower than the baseline.

Table 8. Median, Inter-Quartile Range, and 25th Percentile of Conflict Warning Time

Run	GAB	CGS	Lon TM	Lon CD	Med	IQR	25th %
1	Off	Off	0.50	1.00	412.0	251.0	299.0
2	Off	Off	0.50	1.25	417.0	288.5	305.5
3	Off	Off	0.75	1.50	426.0	300.0	319.0
4	Off	Off	1.00	1.00	412.0	243.0	304.0
5	Off	On	0.50	1.50	322.0	203.0	259.0
6	Off	On	0.75	1.00	318.0	189.5	256.5
7	Off	On	1.00	1.25	319.0	196.0	256.0
8	Off	On	1.00	1.50	308.5	177.0	257.0
9	On	Off	0.50	1.50	418.0	261.0	309.0
10	On	Off	0.75	1.00	416.0	246.5	316.0
11	On	Off	1.00	1.50	415.5	237.5	310.5
12	On	On	0.50	1.00	308.5	190.5	258.0
13	On	On	0.50	1.50	315.0	195.5	257.0
14	On	On	0.75	1.25	313.0	194.0	257.5
15	On	On	1.00	1.00	310.0	177.0	255.0
16	On	On	1.00	1.50	318.0	189.0	258.0
17	Off	Off	1.25	1.25	418.0	251.0	315.0
18	Off	On	1.25	1.25	329.0	193.5	257.0
19	On	Off	1.25	1.25	416.0	238.5	314.0
20	On	On	1.25	1.25	326.0	190.5	257.5
BL	Off	Off	1.50	1.50	490.0	464.5	342.5
IBL	Off	Off	1.50	1.50	471.0	412.0	319.0

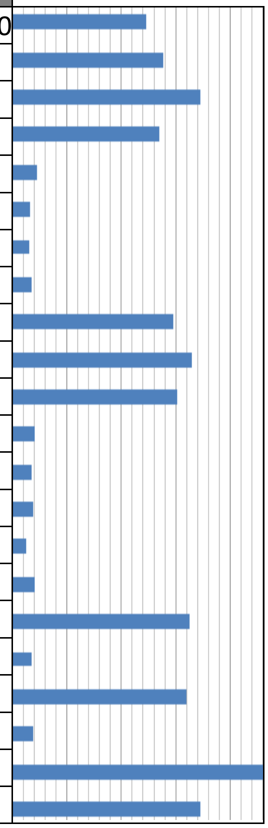

Figure 7 shows the Hit Rate [Crowell et al, December 2011a] versus the FA Rate for each of the 20 treatment runs and the BL. Each point on the plot represents one of the 20 treatment runs or the BL and is labeled as such. The color of each point represents the percentage difference of the 25th percentile of warning time from the BL. The legend to the right of the plot shows that the color moves from blue to red as the warning time decreases. Since the goal is to not decrease warning time, a blue color is desirable. The goal is also to decrease FA Rate while retaining Hit Rate, so the most desirable location is the top-left corner of the plot.

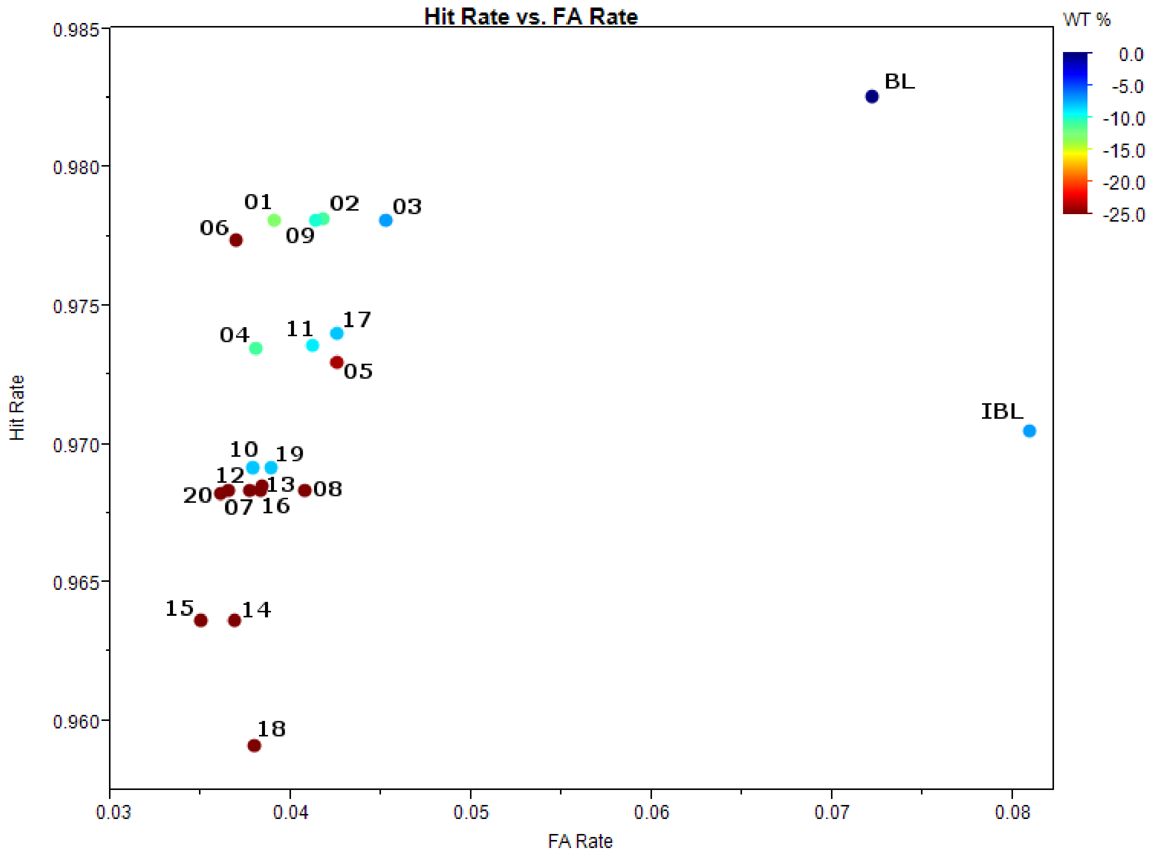

Figure 7. Hit Rate vs. False Alert Rate (Colored by Warning Time)

There is no one run that stands out in the top-left corner of this plot. Instead, there are a group of runs that perform relatively well in Hit Rate and FA Rate, but none of which perform as well as the BL in Hit Rate. All of the treatment runs, however, greatly surpass the performance of the BL for FA Rate. None of the treatment runs perform as well as the BL for Warning Time. A dark blue color is desirable, but Run 3 provides the best with a blue color only slightly darker than the others.

Because of the smaller search space represented by this experiment, the performance of the treatment runs are packed tightly between 0.03 and 0.05 FA Rate. There does, however, seem to be two groups of runs in the warning time. Nine of the runs have warning time colors in the blues and greens, whereas the other eleven are all dark red. The common setting of each of these runs is the CGS prototype. All of the red runs have CGS turned on, whereas all of the blue and green runs have it off.

The CGS prototype is expected to lower the warning times, since the prototype works by delaying the notification of alerts, and it has shown this in previous experiments, but this is the first experiment in which the CGS prototype creates two obvious groups of warning time performance. However, as shown in Table 8, even with the 25% reduction in warning time observed here, at a minimum value of 255 seconds, the 25th percentile of warning time is still well above the 180 second requirement.

3.1.1 Direct Comparisons

Several runs in the integrated experiment were strategically chosen to provide the capability of performing direct comparisons between the two runs with only a single prototype differing between them. Two sets of runs were generated for each prototype for the purpose of direct comparison.

3.1.1.1 Growth Adherence Bounds

Four sets of runs are available for direct comparison of the GAB prototype. All of the runs exhibit an improvement to FA performance. Two of the runs exhibit an improvement to LA count, whereas the two others show degradations. One run shows an improvement to MA performance, two show degradations, and one has no change.

Table 9 illustrates the set of directly comparable runs for GAB. For FA and LA+MA the counts are shown for when the prototype is off and when it is on, and similarly, the WT values are shown when GAB is off and on. The "Diff" value provided for each metric is the following formula:

$$D = \frac{x_{on} - x_{off}}{x_{off}}$$

where x_{on} is the value of the metric (FA count, LA+MA count, WT) when the prototype is on, x_{off} is the value of the metric when the prototype is off, and D is the resulting "Diff" metric. For the FA and LA+MA metrics, it is desirable to decrease the value with the prototype, so a negative value is most desirable. For the WT metric, it is desirable to increase the value, so a positive value is most desirable.

This table shows clearly that GAB has a significant impact on FA performance, reducing the counts by 5-10%. It is also clear that GAB has no impact on WT and minimal to no impact on LA+MA.

Table 9. Results from Directly Comparable Runs for GAB

			FA			LA+MA			WT		
CGS	Lon TM	Lon CD	Off	On	Diff	Off	On	Diff	Off	On	Diff
On	0.5	1.5	715	644	-0.099	6	7	0.167	259	257	-0.008
On	1	1.5	685	643	-0.061	7	7	0.000	257	258	0.004
On	1.25	1.25	638	606	-0.050	9	7	-0.222	257	257.5	0.002
Off	1.25	1.25	714	653	-0.085	6	7	0.167	315	314	-0.003

3.1.1.2 Conflict Geometry Separation

There are five sets of runs for direct comparison of the CGS prototype. All of the runs improve the performance of the probe for FA count. All runs also degrade the performance of the probe for LA count significantly. Two runs improve the performance on MA count, but only by a single alert each.

Table 10 shows the set of directly comparable runs for CGS. For FA, LA, and MA, the counts are provided for when the prototype is off and when it is on. The "Diff" value provided for each metric is the following formula:

$$D=\frac{x_{on}-x_{off}}{x_{off}}$$

where x_{on} is the value of the metric (FA, LA, MA count) when the prototype is on, x_{off} is the value of the metric when the prototype is off, and *D* is the resulting "Diff" metric. For each of the metrics used, it is desirable to decrease the value with the prototype, so a negative value is most desirable.

In this experiment, unlike the previous experiments, the CGS prototype has the parallel algorithm turned off in all runs. Table 10 clearly shows that there is not nearly as large of a negative impact on LA and MA performance as there was in previous experiments. Though some of the runs do have statistically significant increases to LA+MA, others have an insignificant increase or no change at all. On the other hand, all runs indicate at least a 16% decrease in warning time.

Table 10. Results from Directly Comparable Runs for CGS

			FA			LA+MA			WT		
GAB	Lon TM	Lon CD	Off	On	Diff	Off	On	Diff	Off	On	Diff
On	0.5	1.5	694	644	-0.072	5	7	0.400	309	257	-0.168
On	1	1.5	692	643	-0.071	6	7	0.167	311	258	-0.169
On	1.25	1.25	653	606	-0.072	7	7	0.000	314	257.5	-0.180
Off	1.25	1.25	714	638	-0.106	6	9	0.500	315	257	-0.184

Of particular interest is the third run in the table, which has the parameter and prototype settings that have been recommended by previous experiments. In this run, CGS shows no increase in the LA+MA, and a 7% decrease in FA. It does have an 18% decrease in WT, but once again, it still results in a WT well above the 180 second requirement. With these settings, the total improvements over the current live system are 55.5% reduction to FA, 19% reduction to WT, and no change to LA or MA. Although there is an increase to the LAs and MAs observed from the BL scenario, the BL includes the FA32 enhancements that are not currently in the live system. Combined with these FA32 enhancements, the LA and MA performances balance out to the level of the IBL scenario.

3.1.2 Model Analysis

The prediction profiler[1] is used to examine the results of the integrated model. Since the metrics used in the prediction profile are the percentage difference from the baseline, when the model is set to the baseline settings all metrics are expected to be close to zero. In this specific integrated model, all treatment runs used a Lateral setting of 1, while the baseline used for comparison of False Alerts, Late Alerts, and Warning Time used a Lateral setting of 2.5 nm. Since the baseline was not included in the model, it cannot be shown as a prediction profiler figure.

As this experiment is a multi-dimensional problem (12 dimensions total) it cannot be fully described with a single two-dimensional figure. Figures 8 through 17 attempt to describe the shape of this 12-dimensional hypercube.

1 The prediction profiler is an analysis tool provided by the JMP® software and is described in detail in the Experiment 1 document [Crowell et al, December 2011a].

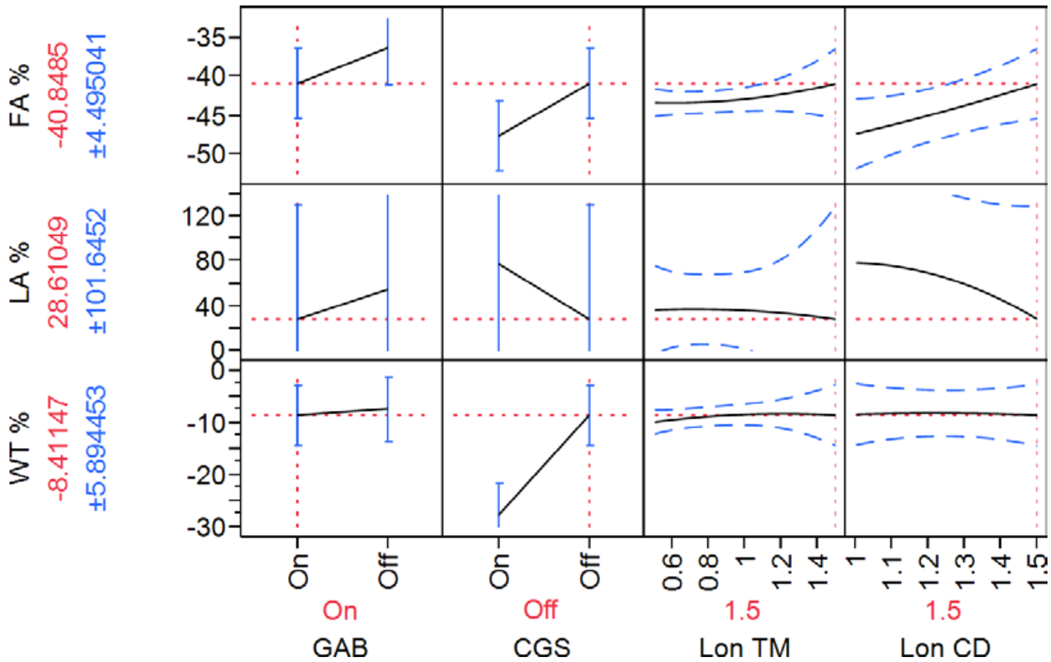

Figure 8. Model Results with Baseline Settings, GAB enabled

Figure 8 shows the model results when all parameters are left at the baseline settings (Lon TM at 1.5 and Lon CD at 1.5) and only the GAB prototype is enabled. The major impact is in the False Alerts which are decreased by 41%. That is a statistically significant impact, and represents a decrease of 503 False Alerts. Late Alerts are increased by 28%, but there are only 4 Late Alerts in the Baseline Scenario, so this represents an increase of 1 late alert which is not statistically significant. It is clear that the GAB prototype alone provides a great benefit to False Alert rate without negatively affecting the other metrics, but it is possible that it contains interactions with other factors.

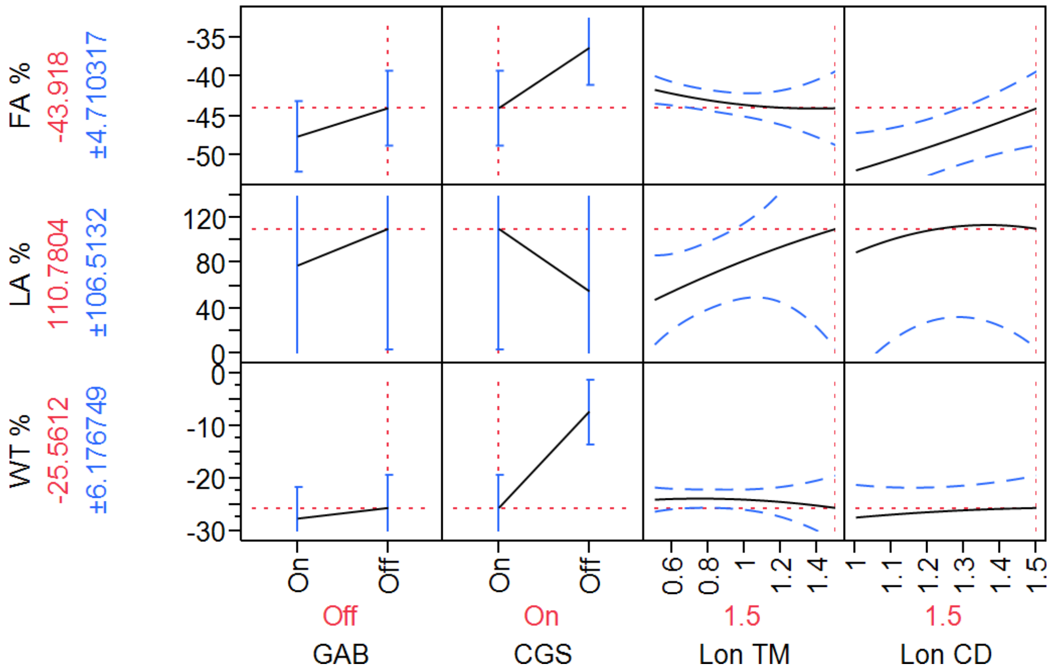

Figure 9. Model Results with Baseline Settings, CGS enabled

In Figure 9 the model results are shown once again at baseline settings, but this time with only the CGS prototype enabled. The major impact is in the Warning Time, which is decreased by 26%, which represents a reduction of about 88 seconds. This is still above the 180 second minimum requirement. Late Alerts are doubled (from 4 to 8), but since the count is so low this is extremely noisy and very little can be inferred from these results. False alerts also decrease by 44%, which is a similar effect to having GAB enabled, but GAB has a much weaker impact on Warning Time and Late Alerts.

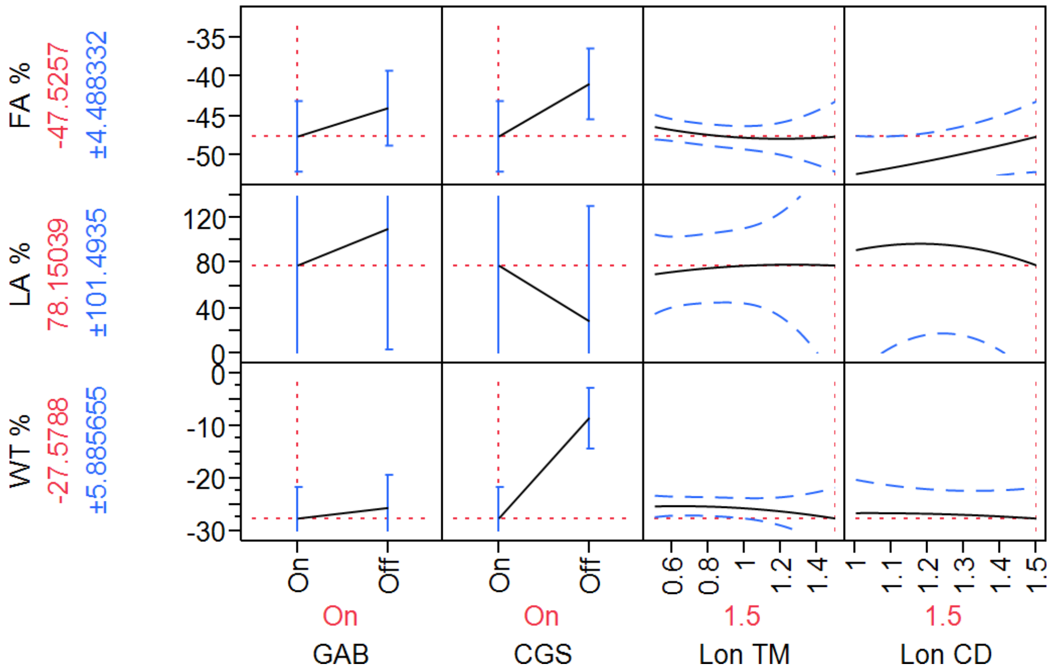

Figure 10. Model Results with Baseline Settings, GAB, and CGS enabled

Figure 10 illustrates the case in which both of the prototypes are enabled and with the other parameters at baseline settings. False Alerts are reduced by 48% (584 alerts) and while the Late Alerts are increased by 78%, though this increase corresponds to a little over 3 LAs and is highly subject to noise effects. Warning Time decreases by 28%, or about 96 seconds. In looking at the second column, it's evident that the CGS prototype has the largest effect on each metric, while GAB has a significant but smaller effect. The LonTM and LonCD have very little effect on LA and WT, and a modest effect on FA.

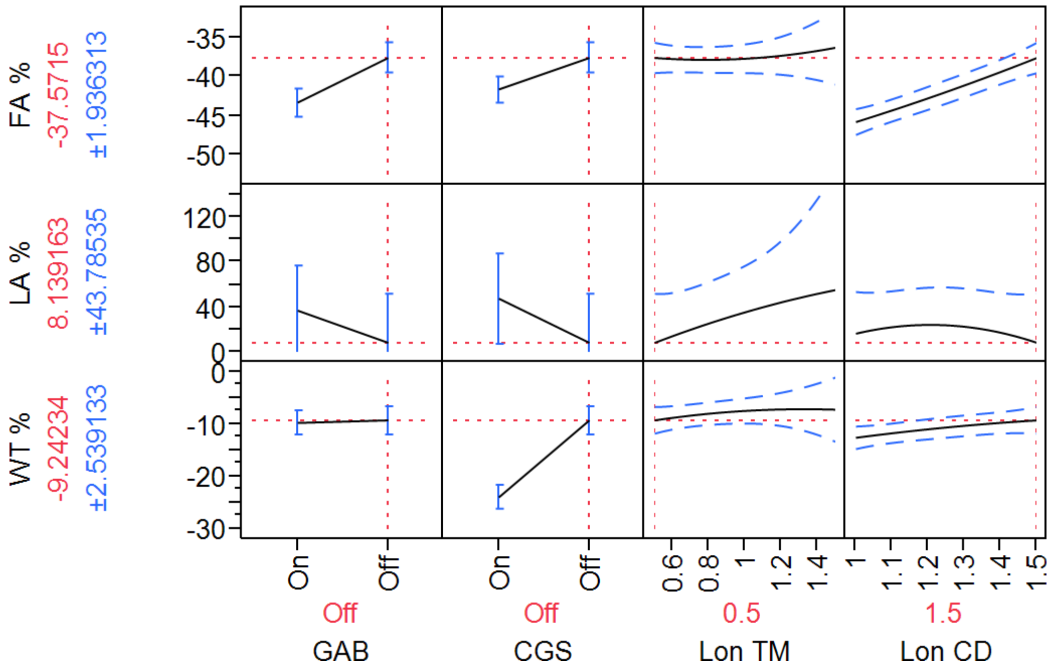

Figure 11. Model Results with Lon TM at .5, GAB, and CGS disabled.

Figure 11 illustrates the case in which both of the prototypes are disabled, Lon TM is set at .5, and Lon CD is set at the baseline level. At these settings, False Alerts are reduced by 38% (457 alerts), a significant result. LAs are increased by 8% (less than 1 flight) and WT is reduced by 9 % (31 seconds), both of which are of minimal significance. In looking at the third column, it's evident that changing the Lon TM setting will have only a small effect on FAs and WT but a sizable effect on LAs, though again it should be noted that there is a large amount of noise in the modeling of LAs due to the low LA count.

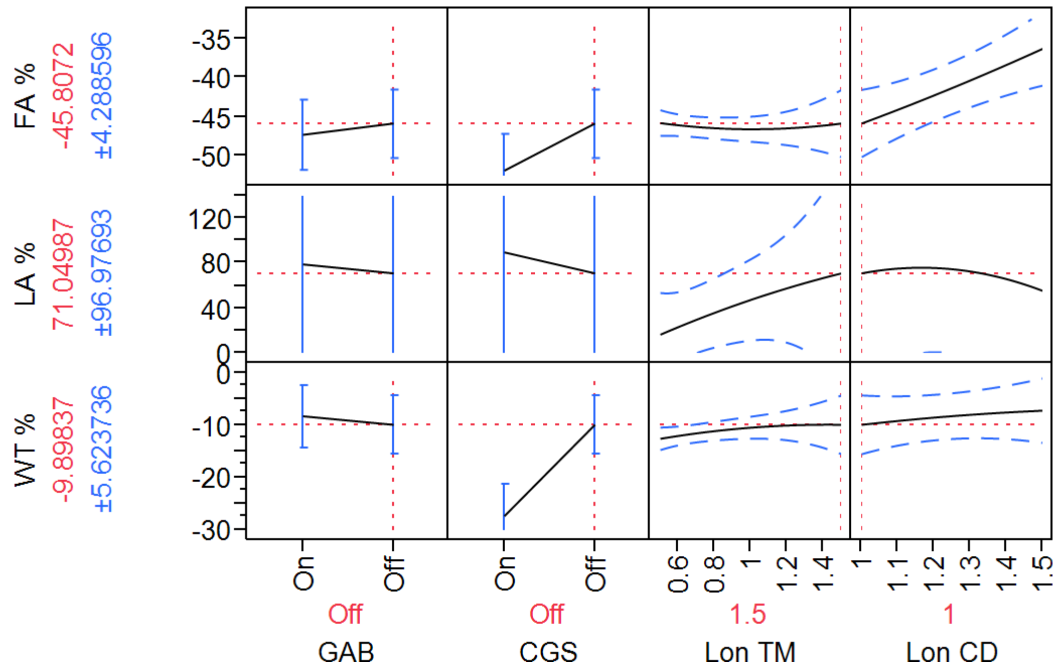

Figure 12. Model Results with Lon TM at .5, GAB, and CGS disabled.

Figure 12 depicts the case in which both of the prototypes are disabled, Lon TM is at the baseline level, and Lon CD is set at 1. At these settings, FAs are reduced by 46% (564 alerts), a significant result. LAs are increased by 71% (about 3 flights) and WT is reduced by 10 % (34 seconds). The percentage change LAs is quite high, while the reduction in WT is not significant. In looking at the fourth column, one can see that changing the Lon CD setting substantially alters FA% but has very little effect on LA% and WT%.

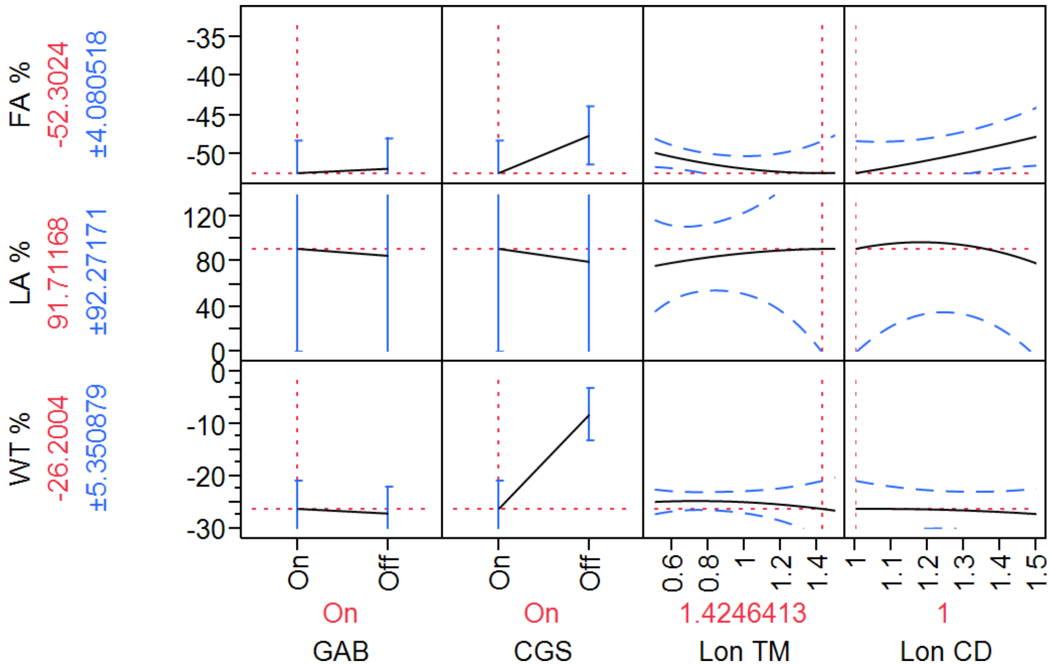

Figure 13. Model Results for Optimal False Alert performance

In Figure 13, the settings indicated minimize the amount of FAs while ignoring the other two metrics. GAB and CGS are both enabled, Lon TM is set at 1.42 which is very close to the baseline setting of 1.5, and Lon CD is set at 1. This results in a 52% reduction in FAs (about 633 flights). CGS and Lon CD have the biggest effect on FA%, and GAB has very little effect. The tradeoff for minimizing FAs is that LAs increase by 92% (about 4 flights) and WT decreases by about 26 % (89 seconds).

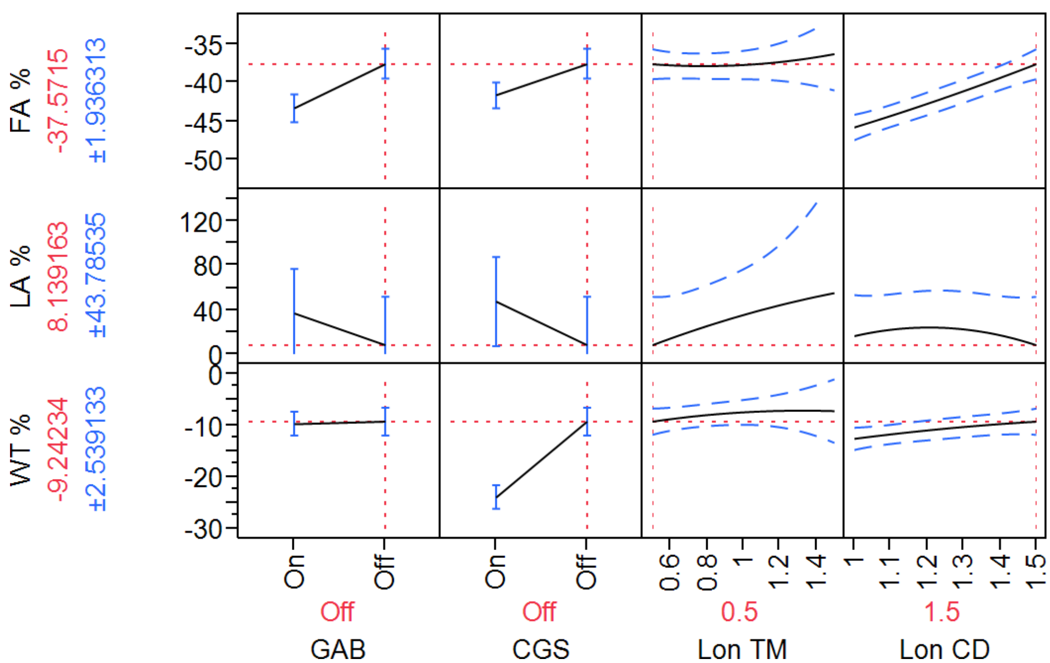

Figure 14. Model Results for Optimal Late Alert performance.

In Figure 14, the settings indicated minimize the amount of LAs while ignoring the other two metrics. GAB and CGS are both disabled, Lon TM is set at the baseline setting of 1.5 and Lon CD is set at .5. This results in only an 8% increase in LAs (less than 1 flight). GAB, CGS, and Lon TM all have a sizable effect on LA% while Lon CD has a minimal effect. The tradeoff for minimizing LAs is that FAs are only decreased by 38% (462 flights). However, WT only decreases by about 9 % (31 seconds).

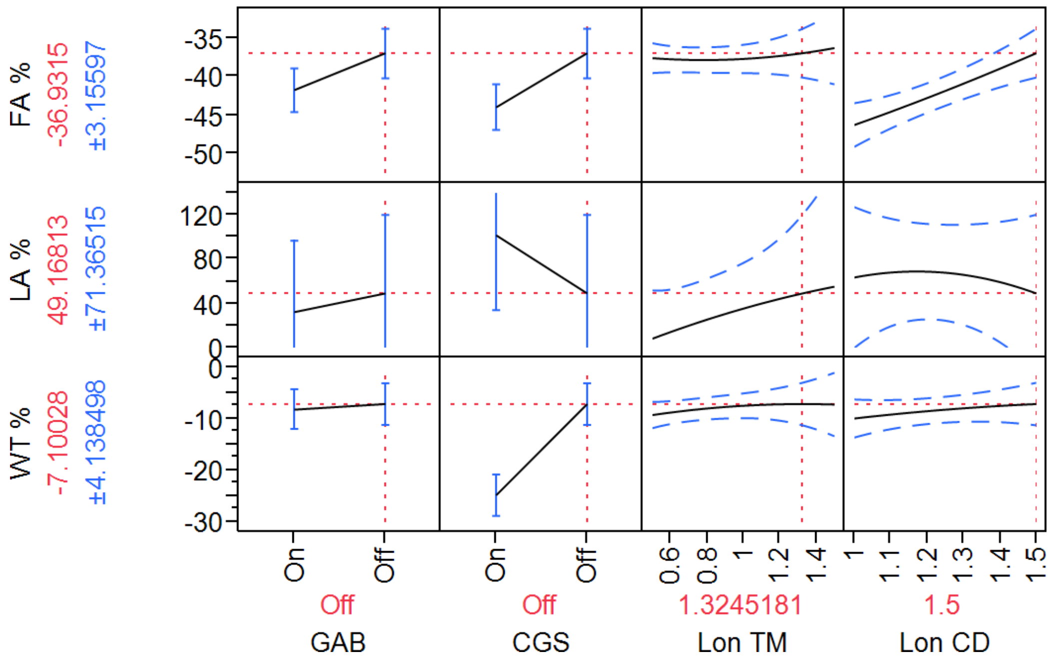

Figure 15. Model Results for Optimal Warning Time performance.

In Figure 15, the settings depicted maximize the amount of WT while ignoring the other two metrics. GAB and CGS are both disabled, Lon TM is set at 1.32 and Lon CD is at the baseline setting of 1.5. This results in a 7% decrease in WT (24 seconds), which is very similar to the value obtained when optimizing LA performance. CGS impacts WT much more than any other setting. However, the tradeoff for optimizing WT is a 49% increase in LAs (about 2 flights) and a 37% decrease in FAs (450 flights).

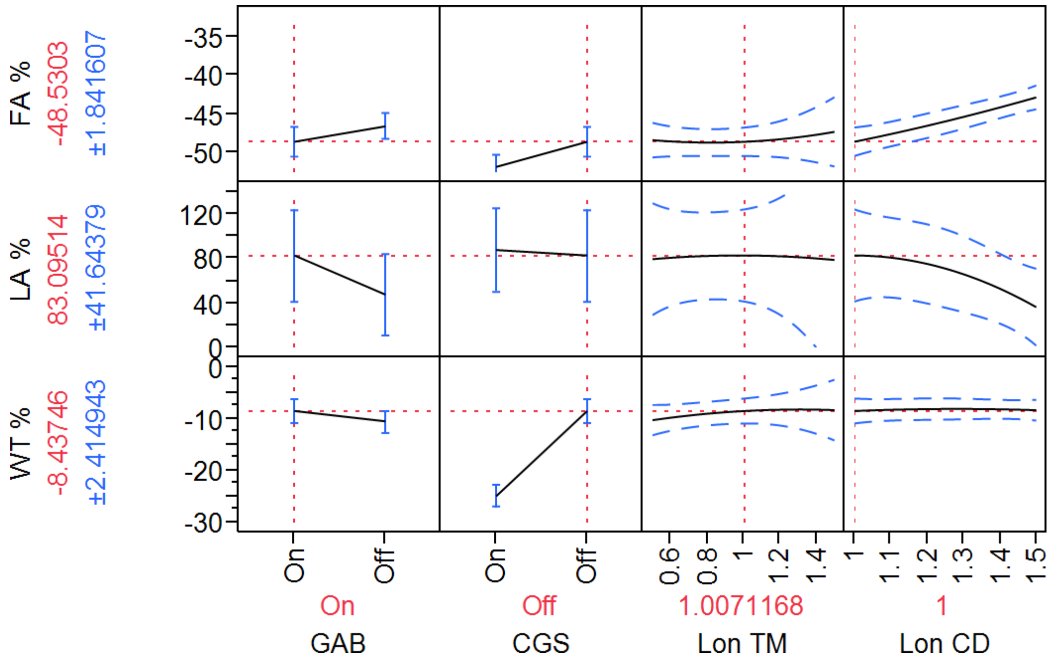

Figure 16. Model Results for Minimizing False Alerts and Maximize Warning Time.

Figure 16 illustrates the settings that provide a balance between FA and WT performance while not considering LA performance due to its high noise levels. GAB is enabled, CGS is disabled, Lon TM is at a middle value of about 1.01 and Lon CD is set to its minimum value of 1. This results in a 49% reduction in FAs (about 596 flights) with only an 8% decrease in WT (about 27 seconds). While LAs increase by 83%, again it should be noted that this is an increase of about 3 flights.

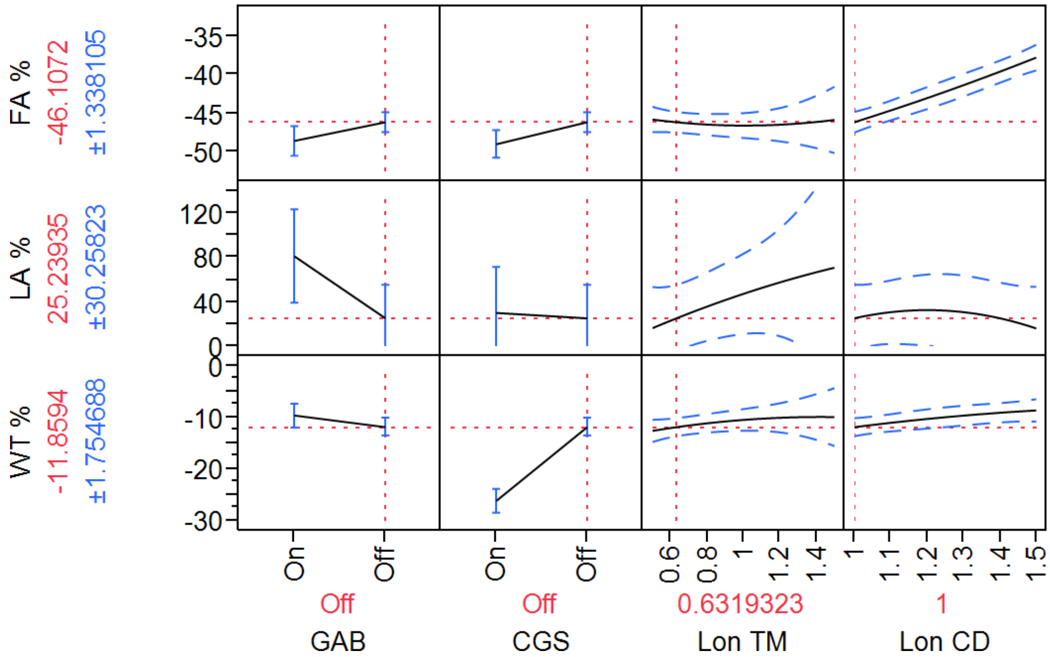

Figure 17. Model Results for Optimizing all 3 metrics.

Figure 17 depicts the model settings that result when attempting to optimize all 3 metrics. FAs and LAs are minimized while WT is maximized. To obtain this result, GAB and CGS are both disabled while Lon TM is set at a low value .63 and Lon CD is set to its minimum value of 1. This results in a 46% reduction in FAs (about 560 flights) and a 12% reduction in WT (about 41 seconds) while only increasing LAs by 25% (about 1 flight).

3.1.3 Factor Effects

The analyses performed in the previous sections have helped to determine the effects of the two prototype enhancements and parameter settings. Conclusions and recommendations can now be made based on this experiment and the three previous experiments. This section describes the effects of each factor on each response in the model.

The Growth Adherence Bound (GAB) prototype displayed a significant improvement to False Alert performance. It can provide up to a 6% reduction in FAs and always provides at least a small amount of improvement. For Late Alerts, up to a 5% reduction and up to a 63% increase, but because of the small sample size, the results of the model are not statistically significant. GAB may reduce the warning time by up to 2%, but can also increase the warning time by the same amount.

The Conflict Geometry Separation (CGS) also displays a significant improvement to FA performance. It also provides up to a 6% reduction in FAs, and at least a 1% reduction. Unlike the results of previous experiments, this version of CGS did not significantly impact LAs and MAs in most cases. It can create up to a 47% increase in LAs and MAs, but at most settings does not generate a statistically significant increase. Of all the factors, CGS had the largest impact on the warning time, reducing it by up to 18%. However, even with an 18% reduction in warning time, the 25th percentile of warning time is still more than a minute above the required 180 second warning time.

The longitudinal track monitoring bound (TM) showed very little significant impact to any of the three factors. However, in some circumstances, setting this parameter below 1.25 nm may increase the FAs by as much as 3%.

The longitudinal conflict detection bound (CD) had a much more significant impact on the performance of the probe. Reducing the parameter from its current setting of 1.5 nm to a setting of 1.0 nm can reduce the FAs by 9% while having no significant impact on LAs and only reducing the WT by 3%.

4 Recommendations and Future Work

This experiment expands upon the analysis done in three previous experiments, by focusing in on certain factors and applying the experiment to a third scenario. This is the last experiment to be performed in this series, so final conclusions and recommendations can now be made. Prior to this experiment, the recommendations were to set the lateral conformance bound to 1.0 nm and the longitudinal to 1.25 nm with Growth Adherence Bounds (GAB) and Conflict Geometry Separation (CGS) enabled. The recommendation for lateral conformance bounds was confident enough to be considered a final recommendation, but the others needed to be confirmed. This experiment attempted to confirm the GAB and CGS recommendations and to expand upon the longitudinal recommendation to see if a little more performance could be squeezed out of it.

In this experiment, the GAB prototype showed as much as a 10% improvement to False Alerts (FA) without any impact on Late Alerts (LA), Missed Alerts (MA), or warning time (WT). This is consistent with previous experiments, and as a result, it is the final recommendation of the Concept Analysis Branch that this prototype be pursued for addition into the live ERAM system.

The CGS prototype also showed as much as a 10% reduction in FAs. It did, however, have a small impact on LAs and MAs, increasing them by up to 50%. It is important to note, however, that, because of the small sample size, a 50% increase represents only two to three alerts. Although the CGS prototype on its own may increase the LAs and MAs, combined the with FA32 enhancements, it shows no impact to the LAs and MAs at the recommended settings. The major negative impact of the CGS prototype is now the WT. It may reduce the WT by up to 19%. However, even with this 19% reduction, the WT is still well above the 180 second requirement. Though CGS had to undergo many modifications to reach its current performance levels, the results observed here are consistent with previous experiments, and so it is the final recommendation of the Concept Analysis Branch that this prototype be pursued for addition into the live ERAM system.

For this experiment, the longitudinal conformance bound was split into its two parts: track monitoring (TM) and conflict detection (CD). This was done in order to determine if any additional performance gains could be observed by modifying the two separately. It was determined that modifying the TM bound by itself has almost no impact to any of the factors, though setting it lower than the current recommendation of 1.25 nm may actually negatively impact performance. The CD bounds have a much greater impact to the probe, with the possibility of reducing the FAs by up to 9% with no significant impact to LAs, MAs, or WT. Once again, these findings are consistent with previous experiments, so it is the final recommendation of the Concept Analysis Branch that the longitudinal conformance bounds be set to 1.25 nm for the TM and CD portions.

Table 11. Final recommendations for the settings and prototypes of ERAM

Prototypes	Lateral	Longitudinal	Likelihood
FA32 ap1; GAB; CGS	1.0	1.25	4|8|20

Table 11 shows the final recommendations being made for all prototypes and parameter changes studied in the five experiments [Crowell et al, December 2011a] [Crowell et al, December 2011b] [Crowell et al, December 2012a] [Crowell et al, December 2012b]. This table includes the recommendations of the likelihood experiment being performed concurrently. With these settings, for this scenario, the False Alerts were reduced by 55.5% and 25th percentile of warning time was reduced by 19%. Missed Alerts and Late Alerts were not affected at all.

5 List of Acronyms and Abbreviations

32BL	FA**32 B**ase**l**ine
AJE-15	FAA Domain Engineering Group
ANG-C41	FAA Concept Analysis Branch
ANSP	**A**ir **N**avigation **S**ervice **P**rovider
ARTCC	**A**ir **R**oute **T**raffic **C**ontrol **C**enter
ATC	**A**ir **T**raffic **C**ontrol
ATO-E	Air Traffic Organization En Route Program Office
BL	FA32 **B**ase**l**ine
CGS	**C**onflict **G**eometric **S**eparation
CP	**C**onflict **P**robe
DST	**D**ecision **S**upport **T**ool
ERAM	**E**n **R**oute **A**utomation **M**odernization
FA	**F**alse **A**lert
FA18	**F**unction **A**rea **18**
FA32	**F**unction **A**rea **32**
FAA	**F**ederal **A**viation **A**dministration
FAR	**F**alse **A**lert **R**ate
FTR	**F**orced **T**rajectory **R**ebuild
GAB	**G**rowth **A**dherence **B**ounds
Horz	**Hor**i**z**ontal
IBL	**I**nitial **B**aseline
IQR	**I**nter-**q**uartile **R**ange
JPDO	**J**oint **P**lanning and **D**evelopment **O**ffice
LA	**L**ate **A**lert
LAR	**L**ate **A**lert **R**ate
Lat	**Lat**eral
Llh	**L**ike**l**i**h**ood
LM	**L**ockheed **M**artin Corporation
Long	**Long**itudinal
MA	**M**issed **A**lert
MITRE	The **MITRE** Corporation
NAS	**N**ational **A**irspace **S**ystem
NC	Correct **n**o-**c**all
NextGen	**Next Gen**eration Air Transportation System
nm	**N**autical **m**iles
SME	**S**ubject **M**atter **E**xpert
TBO	**T**rajectory **B**ased **O**perations
TM	**T**rajectory **M**odeling
TRACON	**T**erminal **R**adar **A**pproach **Con**trol Center
UTC	**C**oordinated **U**niversal **T**ime
VA	**V**alid **A**lert
Vert	**Vert**ical
VHF	**V**ery **H**igh **F**requency
WT	**W**arning **T**ime

References

Bolczak, R., J. Celio, M. Exum, K. Viets. September 2010. "Separation Management Technology Transfer," Coordination Draft, Rev. 1, MP100138R1, The MITRE Corporation, CAASD, McLean, VA.

Crowell, Andrew, Andrew Fabian, Christina Young, Ph. D., Ben Musialek, Mike Paglione. December 2011a. "Evaluation of Parameter Adjustments to the En Route Automation Modernization's Conflict Probe," FAA Technical Note, DOT/FAA/TC-TN12/2.

Crowell, Andrew, Andrew Fabian, Christina Young, Ph. D., Ben Musialek, Mike Paglione. December 2011b. "Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe," FAA Technical Note, DOT/FAA/TC-TN12/3.

Crowell, Andrew, Andrew Fabian, Christina Young, Ph. D., Ben Musialek, Mike Paglione. December 2012a. "Evaluation of Prototype Enhancements to the En Route Automation Modernization's Conflict Probe with an Updated Traffic Sample from Chicago," FAA Technical Note, DOT/FAA/TC-TN12/56.

Crowell, Andrew, Brian Schnitzer. December 2012b. "Analysis of Likelihood Parameter Changes in the En Route Automation Modernization," FAA Draft Technical Note.

Federal Aviation Administration. December 2004. "Next Generation Air Transportation System Integrated Plan," Joint Planning and Development Office, 1500 K Street, Suite 500, Washington, DC 20005.

Federal Aviation Administration. June 2007. "Concept of Operations for the Next Generation Air Transportation System," Version 2.0, Joint Planning and Development Office, 1500 K Street, Suite 500, Washington, DC 20005.

Lapihuska, Charlie. February 2011. "FA18 Reduced Strategic Alerts Prototype Summary Interim 2," Lockheed Martin Report, Separation Management Modern Procedures.

Lapihuska, Charlie. November 2011. "ERAM Conflict Probe Enhancement: Prototype of Reduced Nuisance Alert Options," Lockheed Martin Report, Separation Management Modern Procedures.

McKay, Edward, Ph. D. February 2011. "ERAM Trajectory Lateral Modeling Enhancements, Prototype Effort," Lockheed Martin Report, Separation Management Modern Procedures.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, December 2011.

Oaks, Robert D., Mike M. Paglione. November 2001. "Generation of Realistic Air Traffic Scenarios Based on Recorded Field Data," Air Traffic Control Association (ATCA) 46th Annual Conference Proceedings.

Paglione, Mike M., Robert D. Oaks, Karl D. Bilimoria. November 2003. "Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data," Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Technology, Integration, and Operations (ATIO) Technical Forum.

Rosen, N et al. December 2008. "Analysis of En Route Problem Prediction Operational Use for Separation Management," MTR080389, The MITRE Corporation, CAASD, McLean, VA.

Ryan, Dr. Hollis F., George Chandler, Confesor Santiago, Mike M. Paglione, Shurong Liu. November 2008. "Evaluation of En Route Automation's Trajectory Generation and Strategic Alert Processing: Analysis of ERAM Performance," FAA Technical Note, DOT/FAA/TC-TN08/10.